International Core Journal of Engineering Volume 12 Issue 2, 2026
ISSN: 2414-1895 DOI: 10.6919/ICJE.202602_12(2).0010

Building Surface Defect and Damage Detection Method based on YOLOv8-act

Wenhao Li[a], Huanxin Zhou[b], and Jiaqi Li[c]

School of Civil Engineering, University of Science and Technology Liaoning, Anshan 114051, China

[a]2423473996@qq.com, [b]2483526797@qq.com, [c]lijiaqi@ustl.edu.cn

Abstract

Detecting surface defects and damage on buildings is crucial for ensuring structural safety and durability. Traditional manual inspection methods are inefficient and highly subjective, making them inadequate for large-scale, high-precision engineering demands. This paper proposes a building surface defect detection method based on YOLOv8-act. By replacing the original SILU activation function with the SELU activation function-which possesses self-normalizing properties-the model's ability to extract features from low-contrast, irregularly shaped defects is enhanced. Experimental results demonstrate that the YOLOv8-act model achieves an mAP50 of 0.641, outperforming other compared models. This approach maintains real-time processing and deployment convenience while improving detection accuracy, providing an effective solution for automated and precise detection of building surface defects.

Keywords

YOLOv8-act; Building Surface Defect Detection; Object Detection; Deep Learning; Computer Vision.

1. Introduction

Buildings serve as spaces where people live, work, study, entertain, or engage in other activities. The safety of their structural performance directly impacts the safety of people's lives and property. Over time, under the combined effects of natural elements like wind and sun exposure, as well as human-induced factors such as load application, an increasing number of buildings exhibit defects and damage including cracks, spalling, and corrosion. These defects not only compromise the building's visual integrity and aesthetic appeal but also pose significant risks to structural health and safety. Consequently, the inspection of surface defects and damage in buildings has become a critical requirement in contemporary engineering projects.

Traditional inspection work primarily relies on manual inspection. However, in today's society that constantly pursues efficiency, manual inspection is too inefficient and struggles to meet the demands of large-scale building inspections. Furthermore, inspection results are easily influenced by the inspector's experience and working condition, potentially leading to issues such as missed or misjudged damage, insufficient accuracy, and failure to meet modern engineering requirements for standardized and precise inspection work. However, with advancements in modern computer technology and the application of artificial intelligence, computer vision techniques have emerged as a prominent solution for defect detection. Among these, deep learning-based target defect detection methods achieve automatic recognition of defects and damage in images through training on extensive datasets. This approach significantly outperforms manual inspection in both efficiency and

ISSN: 2414-1895 DOI: 10.6919/ICJE.202602_12(2).0010

accuracy, making it suitable for defect detection tasks of all scales. This technology provides an effective solution for inspection work.

In recent years, the YOLO (You Only Look Once) series of algorithms has emerged as a mainstream single-stage detection technique[1-2]. Leveraging its outstanding real-time performance, detection accuracy, and efficiency, it has demonstrated significant advantages across various object detection tasks and is now widely applied in defect detection for roadways, bridges, and other engineering structures. Zhang Zhiheng introduced the C3k2 _ THK module into the backbone network, which combines partial convolution, heterogeneous kernel selection protocol and SCSA attention mechanism to improve feature extraction ability and reduce computational overhead. The Staged-Slim-Neck module is designed on the neck, which adopts double convolution and dilated convolution at different stages and integrates GMLCA attention to enhance feature representation and reduce computational complexity. The MSDetect module is designed in the detection head to improve the multi-scale detection performance. The detection accuracy of steel surface defects is significantly improved, and an efficient and lightweight steel surface defect detection algorithm ELS-YOLO is proposed[3]. Hongyu Wang et al.proposed an improved rail defect detection algorithm based on YOLO11n. The adaptive downsampling (ADown) module was introduced into the backbone network. The global features were retained by two-dimensional average pooling, and the multi-path convolution and local details were extracted by channel segmentation to avoid the loss of fine texture of small defects. The original SOEP-RFPN-MFM neck network is designed, and the SNI, GSConvE and MFM modules are integrated to realize the dynamic weighted fusion of multi-scale features, break through the bottleneck of low efficiency of small target feature aggregation, and provide a lightweight and high-precision scheme for intelligent detection of rail transit[4]. A defect detection algorithm BFD-YOLO for building facades based on YOLOv7 was proposed by Guofeng Wei et al. The lightweight MobileOne module was used to replace the original ELAN module to reduce the model parameters and improve the reasoning speed. The coordinate attention (CA) module is added to strengthen the key feature extraction to suppress the complex background interference. The SCYLLA-IoU (SIoU) loss function is used to accelerate the convergence of the model and improve the recall rate. The algorithm is trained and verified on 1907 image datasets containing three types of defects : delamination, spalling, and tile shedding. Compared with YOLOv7, the accuracy is significantly improved to meet the real-time detection requirements, and it is stable in complex background and small target defect detection. It provides an efficient and accurate solution for automatic inspection of building facades[5]. Kui Gao et al.proposed an automatic detection framework for cracks and damages in historical buildings based on YOLO series models. Four models of YOLOv5, YOLOv8, YOLOv10 and YOLOv11 were selected to improve YOLOv10: self-calibrated convolution (SCConv) was used to replace the backbone network C2fCIB module to enhance small target and low contrast feature extraction, DySample lightweight upsampling operator was introduced to improve reconstruction quality and efficiency, and three multi-level channel attention (MLCA) modules were integrated in front of the detection head to enhance spatial focusing ability. Experiments show that YOLOv10 has the best performance, significant advantages in multi-target detection and slight damage identification, high positioning accuracy and fast reasoning speed. It provides an efficient and accurate automatic damage detection scheme for the protection of historical architectural heritage. In the future, the accuracy of complex scene detection can be further optimized by combining semantic segmentation network and attention mechanism[6]. However, existing YOLO-based building surface defect detection models exhibit insufficient feature extraction specificity when confronting scenarios with diverse defect types, irregular shapes, and low contrast between minute defects and backgrounds, leaving room for improvement in detection accuracy. This paper proposes a detection method based on YOLOv8-act. By introducing the SELU activation function, it optimizes the network's feature extraction capabilities, enhances the recognition and localization performance for various building surface defects, and improves detection accuracy.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202602_12(2).0010

2. Method

2.1 Computer Vision

Computer vision, as a vital branch of artificial intelligence and deep learning, is a technology that mimics biological visual systems, enabling machines to "see" and "understand" images, videos, and various types of visual data. Its implementation primarily relies on feature extraction and object recognition to extract meaningful information, which is then analyzed to make decisions.

Convolutional Neural Networks (CNNs) represent the most widely adopted structural model in contemporary computer vision[7-8]. CNNs constitute a class of deep learning models incorporating convolutional computations and deep architectures. By simulating the local sensitivity of biological vision through localized receptive fields, CNNs connect neurons only to specific regions of input data, enabling focused extraction of fundamental local features such as edges and textures. The weight sharing mechanism ensures that the same convolutional kernel slides across the entire input space, significantly reducing the model's parameter complexity. Pooling operations, through nonlinear downsampling, reduce the spatial dimensions of the image while enhancing the model's robustness to minor deformations. Leveraging CNN's feature extraction capabilities enables the critical task of extracting meaningful information from images. Unlike traditional manually designed features, CNNs directly process raw visual data inputs. By implicitly learning data patterns during training, they autonomously perform hierarchical feature extraction. This approach eliminates cumbersome manual feature engineering while effectively avoiding cumulative errors inherent in human-designed features, enabling full automation from feature extraction to target classification. The figure below illustrates the basic structure of a CNN.

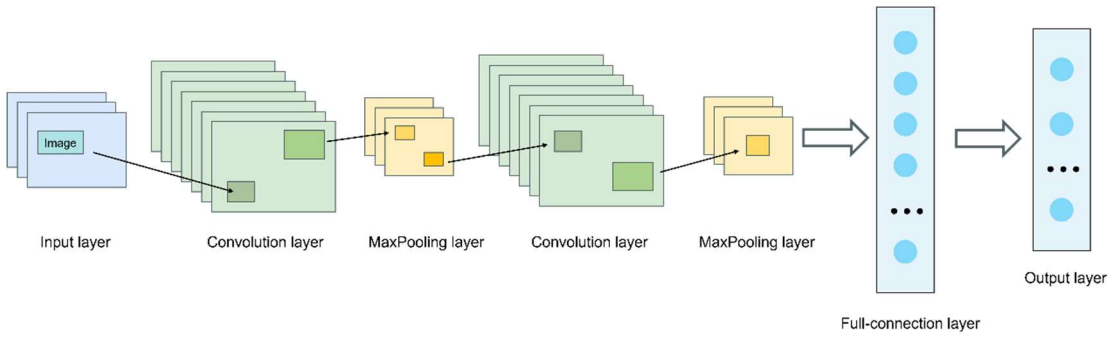


Figure 1. Basic structure of a CNN

2.2 YOLOv8

YOLOv8 is an object detection algorithm released by the Ultralytics team in 2023. It adopts an Anchor-Free architecture[9]. Compared to previous outstanding models in the YOLO series, YOLOv8 represents an advanced and cutting-edge model. It inherits the core strengths of the series-end-to-end detection and high efficiency with precision-while achieving significant improvements in detection performance through multiple structural optimizations. Supporting comprehensive visual AI tasks including detection, segmentation, pose estimation, tracking, and classification, it has become one of the mainstream models for object detection tasks in the field of computer vision.

Taking YOLOv8n as an example, this algorithm comprises four main components: the input layer, backbone layer, neck layer, and head layer [10]. As illustrated in the figure, its backbone network centers on Conv convolution layers, C2f modules, and SPPF spatial pyramid pooling modules. Through multiple rounds of convolution and downsampling operations, it progressively extracts underlying edge and texture features as well as high-level semantic features from the image. Specifically, the convolutional layer performs basic feature extraction and downsampling. The C2f module enhances feature flow through cross-cascading and residual connections, while the SPPF module fuses global information via multi-scale pooling to improve adaptability to defects of varying sizes. The head layer then transforms the feature maps extracted by the backbone into final detection

ISSN: 2414-1895 DOI: 10.6919/ICJE.202602_12(2).0010

results. Through upscaling and feature concatenation, the head layer combines feature maps at different scales to enhance detection capabilities for multi-sized objects. Subsequently, the C2f module further optimizes feature flow to improve expressive power. Finally, the Detect module outputs object bounding boxes, categories, and confidence scores from the processed features, achieving multi-scale object detection.

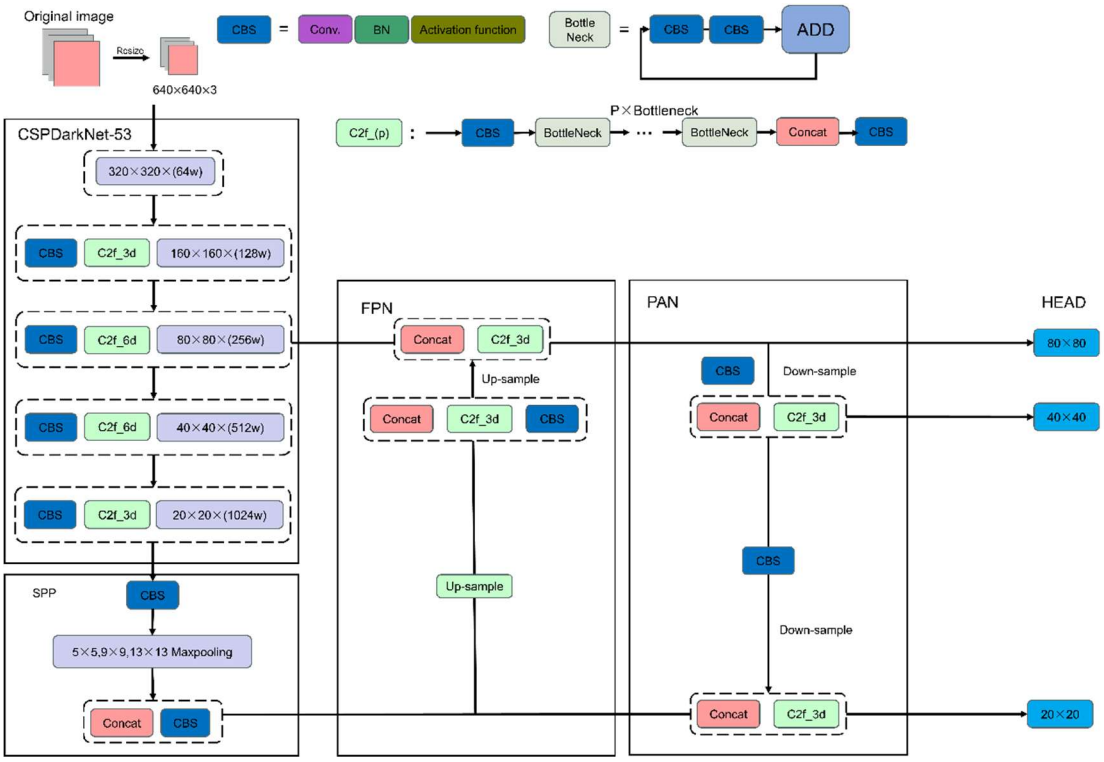


Figure 2. Basic structure of YOLOv8

2.3 YOLOv8-act

Compared to YOLOv8, YOLOv8-act employs the specific SELU activation function to optimize feature extraction. The SELU (Scaled Exponential Linear Unit) activation function possesses self-normalizing properties [11], enabling it to automatically maintain the stability of input data's mean and variance during training. Compared to the SILU activation function commonly used in the original YOLOv8, SELU is defined in its functional form as:

$$SELU(x) = \lambda \begin{cases} x & x > 0 \\ \alpha e^x - \alpha, & x \leq 0 \end{cases}$$

Where:

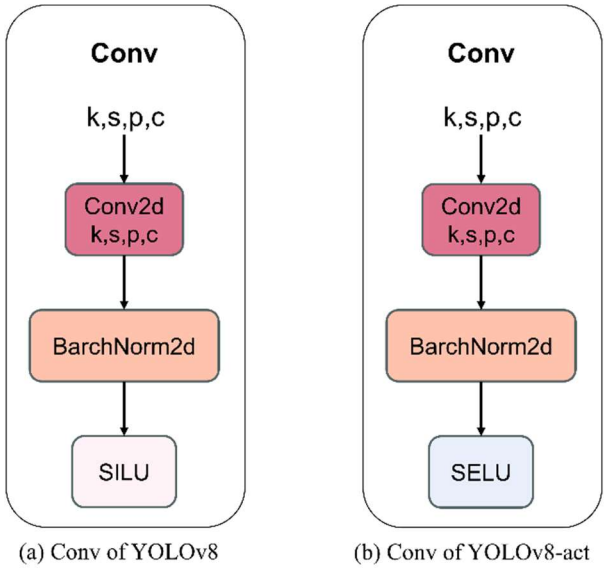
λ refers to the scaling coefficient: it ensures the output of the activation function has zero mean and unit variance, thereby maintaining the stability of the network, and its approximate value is 1.0507,

α denotes the negative half-axis offset coefficient: it determines the curvature of the function when x $\leqslant$0, and ensures the function is continuous and differentiable at x=0, with an approximate value of 1.6733.

This modification is applied to all Conv layers and C2f modules within the model, enabling it to better adapt to the diversity and complexity of surface defects on buildings during feature extraction. Particularly for certain damage types with low contrast against the background or unusual shapes, SELU effectively enhances the model's ability to capture features of surface defects in complex environments without significantly increasing computational load or model complexity. Furthermore,

ISSN: 2414-1895 DOI: 10.6919/ICJE.202602_12(2).0010

YOLOv8-act preserves YOLOv8's original Anchor-Free design, multi-scale feature fusion mechanism, and efficient C2f module. This ensures enhanced detection performance without compromising the model's real-time capabilities and deployment convenience. The figure below illustrates the convolutional layer structures of both models.



(a) Conv of YOLOv8 (b) Conv of YOLOv8-act

Figure 3. Structure Comparison of Conv Modules between YOLOv8 and YOLOv8-act

3. Result

3.1 Data Set Sources and Training Environment

The data used in this study was sourced from the internet and is publicly available on Hugging Face: https://huggingface.co/datasets/xueaidezhouzhou/buildingsurfacedefectdetection/tree/main. This dataset comprises a total of 7,354 images covering six common types of architectural surface defects, including cracks, spalling, exposed reinforcement, delamination, efflorescence, and rust stains[12].

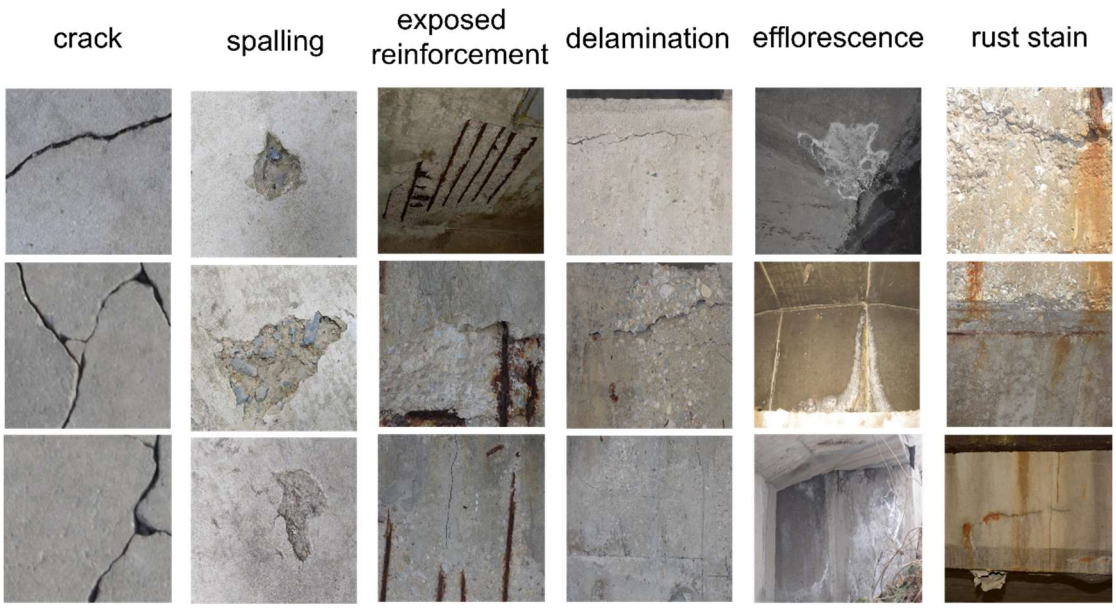


Figure 4. Types of Building Surface Defects

ISSN: 2414-1895 DOI: 10.6919/ICJE.202602_12(2).0010

The model runs on an NVIDIA GeForce RTX 3050Ti Laptop GPU with Windows 10 Pro operating system and CUDA version 11.8. The specific hyperparameter settings for this model are detailed in Table 1, with a maximum training epoch of 100. The batch size is set to 16, the initial learning rate is 0.1%, and the optimizer is configured as SGD[13].

Table 1. Model Hyperparameter Settings

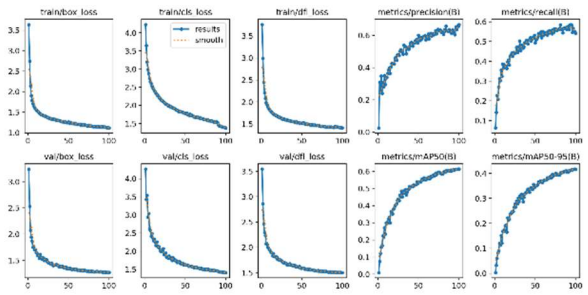
Item	Batch Size	Epoches	Image Size	Optimizer	Initial Learning Rate
Value	16	100	640	SGD	0.01

3.2 Training Metrics Comparison

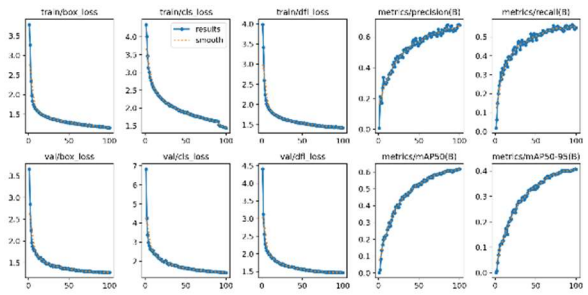
In this study, four models were primarily trained: YOLOv8-act, YOLOv8-AIFI, YOLOv8-ASF, and YOLOv8-ghost-p6. The table below shows the mAP50 values for each model across different defect types. The detection data for various defects in the table reveals that all three models achieved the highest recognition accuracy for Exposed Reinforcement, each exceeding 0.7. Among individual models, YOLOv8-act demonstrated superior performance across most defect categories, particularly in Efflorescence detection. Its mAP50 reached 0.695, significantly outperforming the other three models. This advantage likely stems from the SELU function's exponential behavior in negative regions, which enhances detection of low-contrast textures. In Crack and Spalling detection, YOLOv8-act also achieved the highest value of 0.653, indicating its superior feature capture capability for linear and blocky irregular defects. However, for Rust Stain and Delamination, all three models performed notably poorly, failing to reach even 0.6, suggesting persistent challenges in detecting certain defects.

Table 2. mAP50 Performance Comparison of Different Models on Defects

Defect Type	YOLOv8-AIFI	YOLOv8-ASF	YOLOv8-ghost-p6	YOLOv8-act
Exposed Reinforcement	0.772	0.757	0.738	0.773
Rust Stain	0.561	0.534	0.540	0.561
Crack	0.635	0.620	0.603	0.653
Spalling	0.633	0.640	0.620	0.653
Delamination	0.510	0.498	0.498	0.509
Efflorescence	0.564	0.654	0.580	0.695
Overall	0.612	0.617	0.597	0.641



(a) YOLOv8-AIFI



(b) YOLOv8-ASF

ISSN: 2414-1895 DOI: 10.6919/ICJE.202602_12(2).0010

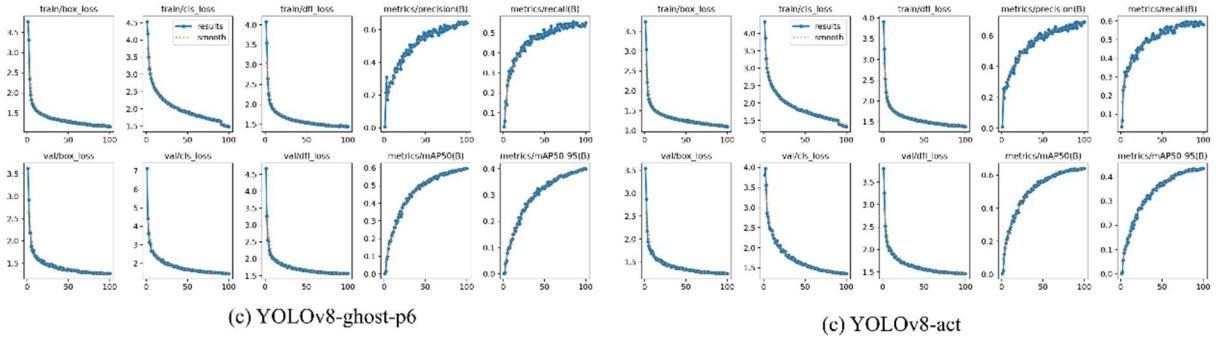


Figure 5. Loss values and mAP for different model

Figure 5 displays the training loss curves and performance metric curves for the four models. The loss curves include box_loss, cls_loss, and df_loss, categorized into training and validation sets: The training loss for all four models rapidly declines from initial high values and stabilizes within 50 epochs, indicating the models efficiently learn defect location features and category features. Simultaneously, the trends and values of validation loss closely align with training loss, without exhibiting scenarios where "training loss drops extremely low while validation loss surges abruptly." This demonstrates strong training generalization capabilities across all models, with no significant overfitting.

Comparing performance metrics across models, YOLOv8-act exhibits significantly higher mAP50 and mAP50-95 convergence values than the other three models. Its precision and recall convergence values are also relatively superior, indicating this model achieves high accuracy while covering more actual defects during training. Conversely, YOLOv8-ghost-p6 exhibits a relatively low mAP50 convergence value, indicating its overall performance is slightly weaker. The figure below shows partial image examples from the YOLOv8-act test.

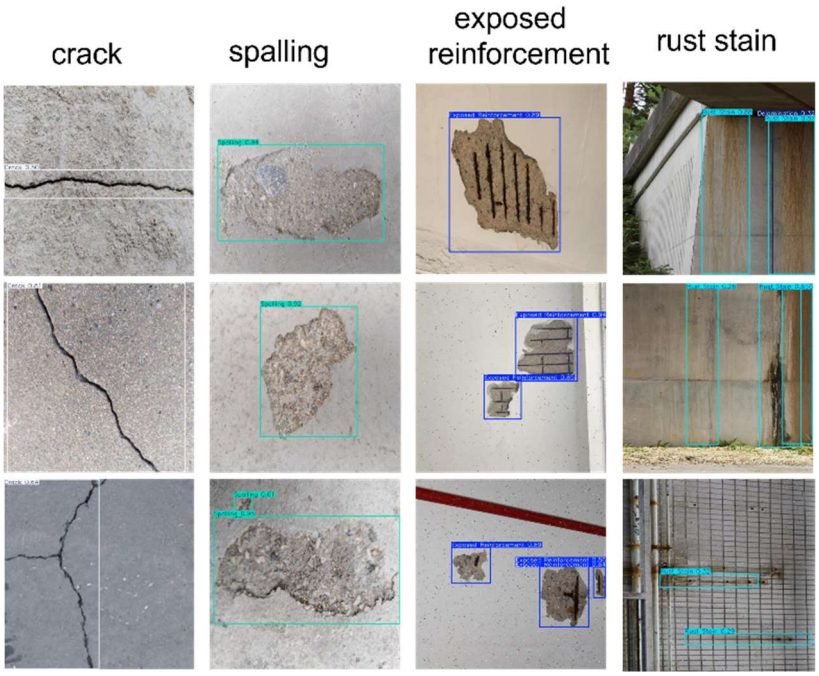


Figure 6. Defect Detection Example

4. Conclusion

This paper proposes a building surface defect detection method based on YOLOv8-act. By replacing the activation function in the YOLOv8 network with the SELU function featuring self-normalization

ISSN: 2414-1895 DOI: 10.6919/ICJE.202602_12(2).0010

properties, the model's ability to extract features from low-contrast, irregular defects is enhanced. Experiments on a public dataset containing six defect categories demonstrate that the YOLOv8-act model achieves an overall detection accuracy (mAP50) of 0.641, outperforming other comparison models. Significant improvements are observed particularly for complex defects such as spalling and efflorescence. This approach provides an effective solution for automated and precise detection of building surface defects. However, the current dataset has limited coverage of defect scenarios and building materials. Future work should expand the sample scope to include defect data under extreme conditions such as low illumination and adverse weather. Additionally, exploring the further integration of SELU with attention mechanisms will enhance the model's generalization capability in complex scenarios.

References

[1] Choutri K, Lagha M, Meshoul S, et al. Fire detection and geo-localization using uav's aerial images and yolo-based models[J]. Applied Sciences, 2023, 13(20): 11548.

[2] Singla R, Sharma S, Sharma S K. Infrared imaging for detection of defects in concrete structures[C]//IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2023, 1289(1): 012064.

[3] Zhang Z, Zhong G, Ding P, et al. ELS-YOLO: efficient lightweight YOLO for steel surface defect detection[J]. Electronics, 2025, 14(19): 3877.

[4] Wang H, Zhao J. Research on Defect Detection on Steel Rails Based on Improved YOLO11n Algorithm[J]. Applied Sciences, 2026, 16(2): 842.

[5] Wei G, Wan F, Zhou W, et al. Bfd-Yolo: A Yolov7-based detection method for building façade defects[J]. Electronics, 2023, 12(17): 3612..

[6] Gao K, Chen L, Li Z, et al. Automated Identification and Analysis of Cracks and Damage in Historical Buildings Using Advanced YOLO-Based Machine Vision Technology[J]. Buildings, 2025, 15(15): 2675.

[7] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.

[8] Oquab M, Bottou L, Laptev I, et al. Learning and transferring mid-level image representations using convolutional neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 1717-1724..

[9] Hussain M. Yolov5, yolov8 and yolov10: The go-to detectors for real-time vision[J]. arXiv preprint arXiv:2407.02988, 2024.

[10] Hussain M. Yolov1 to v8: Unveiling each variant–a comprehensive review of yolo[J]. IEEE access, 2024, 12: 42816-42833.

[11] Klambauer G, Unterthiner T, Mayr A, et al. Self-normalizing neural networks[J]. Advances in neural information processing systems, 2017, 30.

[12] Lin X, Meng Y, Sun L, et al. Building Surface Defect Detection Based on Improved YOLOv8[J]. Buildings, 2025, 15(11): 1865.

[13] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.